SAM: Speech-Aware Applications in Medicine to Support Structured Data Entry

A. K. Wormek, DVM[1], J. Ingenerf, Ph.D.[1], H. F. Orthner, Ph.D.[2]

[1]GSF - National Research Center for Environment and Health, Medis - Institute of Medical Informatics and Health Services Research, Neuherberg, Germany

[2]Department of Medical Informatics, University of Utah, Salt Lake City, Utah, U.S.A.

In the last two years, improvements in speech recognition technology has directed the medical community's interest to porting and using such innovations in clinical systems. The acceptance of speech recognition systems in clinical domains increases with recognition speed, large medical vocabulary, high accuracy, continuous speech recognition, and speaker independence. Although some commercial speech engines approach these requirements, the greatest benefit can be achieved in adapting a speech recognizer to a specific medical application. The goals of our work are first, to develop a speech-aware core component which is able to establish connections to speech recognition engines of different vendors. This is realized in SAM. Second, with applications based on SAM we want to support the physician in his/her routine clinical care activities. Within the STAMP project (STAndardized Multimedia report generator in Pathology), we extend SAM by combining a structured data entry approach with speech recognition technology. Another speech-aware application in the field of Diabetes care is connected to a terminology server. The server delivers a controlled vocabulary which can be used for speech recognition.

INTRODUCTION

Advanced speech technology enables continuous-speech recognition (no pauses between spoken words), features speaker-independence, and provides support for large vocabularies. While in general the accuracy of spoken language recognition systems is not quite satisfactory, it is sufficient for domains with limited complexity. The industry is aggressively pushing ahead with speech recognition products aimed primarily at the transcription market. However, true speech understanding systems are not yet commercially available with the exception of applications with very limited domains. NIST has established evaluation procedures for speech understanding systems but they are limited to reading Wall Street Journal-type Reports and Airline Reservation Tasks. These domains are small compared to the domains represented by internal medicine or family practice and minuscule compared to the domain complexity of the entire health care system.

Within the health care market the several companies offer speech recognition engines with vocabulary and language models in the health care sector. In addition, value added resellers use these voice engines to provide coverage for specific clinical domains. It appears, voice recognition is at the verge of becoming a standard component of workstation and desk-top computers. All major computer operating systems vendors such as Microsoft, IBM, Apple, Sun, ... use voice recognition as a standard feature in their future operating systems and user interfaces. In addition to the commercial research groups, there are a number of university-based and not-for-profit research organizations that have well established speech laboratories. Most of these research laboratories have been established through funding from the Department of Defense Advanced Research Project Agency (ARPA) or the National Science Foundation (NSF).

Speech Recognition Research at the GSF

At first sight, the use of speech-enabled applications in medicine seems to be an attractive way to automate the clinical dictation process. With a more natural user interface, even computer illiterates may find easy access to computer systems. Experience shows, if you confront people with speech recognition, even computer specialists have rather high expectations concerning this novel technology. Implicitly they anticipate the machine react just like a human listener: not only to recognize spoken words but also to understand them. Not only in medicine, acceptance of speech-aware applications seems to require that users also know about the limitations of the current

1091—8280/97/$5.00

technology[1]. Application developers must realize that using speech as an input mechanism is not as simple as adding a new interface on top of an existing text or graphical interface[2]. Even cognitive issues in the potential user's profile play a role in handling speech-aware applications. For example, what are the user's preferences in composing a text, is he more of a linear or a nonlinear thinker[1]? There are two solution to overcome our lack of understanding. One solution is to adapt the speech recognition applications to each user and his/her daily work. The advantage is that the user is not forced to make major changes in his/her work process which increases acceptability by this user. A major drawback of this procedure is that the application supports only one user or at best a limited group of users. Another solution is to pursue a more general approach which focuses on the task to be solved rather than on supporting a specific person or group. To enable speech recognition in this case, the user must be trained and learn how to interact with the speech interface, i.e., learn "when can be spoken what." Thus, the physician must learn the specific components of speech recognition interface. The first solution taxes the system's capacity to adapt to the user and the second solution taxes the users capacity (and tolerance) to adapt and learn. The optimum solution is probable somewhere between these two extreme scenarios and as we gain experience with this technology and its uses we may find principles that can be generalized across users and/or application domains.

Structured Data Entry

There is a long tradition in medical language processing to extract structured data out of narrative free text entries. At least diagnoses and procedures are coded (semi-)automatically with respect to controlled vocabularies[3-5]. But the usability of these natural language processing techniques is still limited[6, 7]. The most problematic point is that there is no „interactive" feedback in the process of data capture.

On the other side, data can also be collected at time of data entry following the concepts of structured data entry (SDE)[8]. The advantages are obvious and clearly stated for example in van Ginnecken[9]. Moorman[10] has outlined requirements that should be fulfilled for acceptability by the physician. SDE involves the use of a predefined structure and vocabulary at the time of data entry. Interface technology such as graphics and voice input may further the efficiency and feasibility of SDE in daily practice[9].

The easiest way to realize SDE is to define relevant entities with their attributes and values and link them to hierarchically organized forms. The terms used for the entities should be linked to a controlled vocabulary to limit the variety of expression. To achieve a more implementation flexibility, the knowledge in the forms are made explicit in a meta-model where the relevant entities are linked to concepts of the controlled vocabulary. The latter is defined by the application and its associated domain knowledge. The user is liberated from the fixed order of data entry and different data entry protocols are possible. However, this approach is restricted with respect to scope and depth of descriptions, as long as there is no mechanism for a compositional concept representation like the one developed in the GALEN project. The GALEN approach provides, on demand via a query to a terminology server, all those expressions that are sensible to say about a concept within the context of specific patient record. Such a dynamic approach requires definitions of composite concepts based on a modest number of primitives and a formalism of concept classification that can compute the concept's semantics with respect to the other concepts[11].

One of the benefits of the GALEN approach is the potential for multilingual report generation[12]. Otherwise report generation is restricted to the use of canned texts that often causes poor and redundant reports. In parallel with the implementation of speech-driven SDE and report generation in the STAMP-project, we started modeling a suitable subset of medical terms in GALEN´s representation language GRAIL as well as linguistic annotations for the NLP group in Geneva. This is part of the GALEN-project for which our institute is responsible within the ongoing health application telematics program of the European Union (EU)[13].

THREE SPEECH-AWARE APPLICATIONS IN MEDICINE

In the following we will present three speech-aware applications in medicine based on SAM. Each application handles speech data input in a specific way, but all conform to the SDE. The first application focuses on speech commands, sometimes called navigation. The second one is a template-based trigger-command oriented system that generates a pathology report. The third application focuses on handling dynamic vocabularies and grammars, which are provided by an authoring tool connected to the GALEN terminology server. Thus, all words and

phrases the speech engine is able to recognize belong to a controlled vocabulary.
One idea of the STAMP project is to combine the benefits of SDE and speech recognition. The user interacts with the system by using both, dictation and control commands of a speech recognition engine. To demonstrate the technical feasibility, we chose a small well-circumscribed domain within pathology: Cytology[14].

The Institute of Clinical Cytology at the Technical University of Munich has been using SDE and automated report generation for a long time. The preventive gynaecological cytological diagnostic service for the national cancer society constitutes a large work load for the institute. Currently, most of the routine examinations are documented on printed forms but they are in the process to establish a computerised medical record system to provide the forms electronically. The cytologists are interested in increasing their reporting efficiency by using speech-driven SDE and report generation in addition to the traditional free text entry. Also, they hope that the new system will guide less experienced cytology assistants in carrying out some examinations.

For the domain of gynaecological cytology, we have carefully defined the entity-attribute-value triplets that are relevant for reporting and canned text generation. For each selection of a valid entity-attribute-value triple a fixed set of general commands available to activate or modify items by spoken input. In addition we allow free text dictation in cases, the cytologist have to describe SDE-recorded data in more detail.

Based on the positive experience in cytology we expanded to the wider field of anatomic pathology. In the „Pathologisches Institut Kempten", the average number of pathologic and microscopic examinations is between 350 and 450 examinations per day. For efficiency reasons, one of the pathologists uses predefined voice templates in conjunction with a Dictaphone transcription system. He dictates commands like 'b1a, stop3, complete, rest as usual' rather than the full description of the microscopic findings or the full-text diagnosis. A trained typist transcribes the dictation substituting the commands with the text of predefined templates. For example, 'b1a' stands for the description of the examination of a skin-excision, no deviation are dictated, so the predefined text in the template can be taken without changes. The command 'stop3' means that the word that follows (i.e., 'complete') has to be inserted into text at the marker 'stop3'. In the final report an expression like „complete excision" after the diagnosis „Verruca vulgaris" will be included. The rest of the template can taken without any change. Afterwards, the pathologist proofreads the transcription.

To support this individual pathologist by a speech recognition system, we analyzed the way he dictates. In the application we have build for him, he is allowed to speak his commands like 'b12' as he is used to do, or he speaks a diagnosis: 'Verruca simplex'. In any case a graphical control window is opened, showing the predefined text template and the text markers. He is able to substitute or supplement text within the markers by using voice commands or dictation. The command: 'generate report' performs the completion of the final report (which includes the demographic patient data, the name of the physician who requested for the examination, etc.). Proofreading and verification is performed during the process of dictation. A related procedure, using text templates with fill-in entries to increase speech recognition performance is described by Meijer[15] and Teplitz[16].

A third speech-aware application incorporates the GALEN terminology server. Our focus here is in capturing data for diabetes patients. With the help of an authoring tool, it is possible for a physician to formulate inquiries to the terminology server for a diabetic concept, which will include all related and sensible modifiers relevant to this concept and patient's context. Based on this responds, the physician selects the items he wants to record. This selection and the data from the terminology server enables the authoring tool to generate dynamically one or more input forms. The same information is also used to enable speech input for each form. Most of commercial speech recognition products support partitioning and control of phrases, also called grammars. A grammar is defined as a structured collection of words and phrases bound together by rules that define the set of all utterances that can be recognized by the speech engine at a given point in time. Phrases can be spoken continuously. Grammars also provide a language model for the speech engine, constraining the valid set of words to be considered, increasing recognition accuracy while minimizing computational requirements. It is possible to switch among different grammars and vocabularies at run time. In a compiled grammar file one can specify in which way embedded silences and mumbles has to be handled by the speech recognition engine. The application developers must define the control mechanisms in the speech-aware application to select appropriate vocabularies and/or grammars. The authoring tool transforms the information mentioned above into a Backus-Naur Form grammar, SAM can

compile and select for a recognition session. Additionally, the authoring tool delivers information to SAM, which words or phrases are defined to be considered as commands and how the application has to perform in case of the occurrence.

A feedback window shows the user the scope of the selected phrases and reminds him what could be said for speech recognition. This is important since for large and complex input forms most user have difficulty remember all possible phrases[17].

IMPLEMENTATION

The vendors of speech recognition engines facilitate the task to develop speech aware applications by providing **S**oftware **D**evelopment **K**its (SDKs) application programming interfaces (APIs). The speech recognizer in SAM, is composed of commercially available speech recognition hardware and software available on PC and 32-bit Windows platforms. We have the most experience with the IBM VoiceType speech recognition system that supports German language. We use IBM-VoiceType system with a pathology vocabulary provided by IBM.

In order to be platform independent, and to stay compatible with implementations in other projects of our working group, SAM and the speech-aware applications use the object oriented programming language VisualWorks Smalltalk. The design of SAM has been guided of Design Patterns[18]. For example, the connection to a speech recognizer is realized in an abstract Smalltalk Class, that decouples the application from the API set of a specific speech recognition engine. In this way, SAM easily adopt to speech recognizers from different vendors. A standardization of the Speech Recognition API would make our effort easier. SAM can be considered as a kernel to provide medical applications an adapter or interface to a speech recognizer. Like an adapter in the background, SAM performs message handling from an application to a speech recognizer, and vice versa. SAM also has control functions. In SAM an API function call establishes the session by linking a speech recognition engine and the application. This causes the recognition engine to shift into a state to receive spoken input and try to recognize it. Recognized utterances are sent back to SAM as a message (ASCII strings). It handles this message and knows from the application what to do next, performing an application command, disable or enable a vocabulary or a grammar, or give an order to the speech engine to listen to the next spoken input. After successful completion of the series of actions that culminate in a transaction, the system returns to the attention listen mode.

ASPECTS OF EVALUATION

Each of the three applications has its own characteristics and its own way to supports the user. An extern detailed evaluation by physicians in their routine clinical care activities, requires an elaborate evaluation concept with concrete assessment principle. More complex than the technical assessment of the recognizer (reliability, word accuracy, actual dictation time, ...) is performance assessment of the application in the whole. Besides human and technical factors (Operating System, processor, sound card, ...), the influence of graphical user interface, the influence of the underlying SDE, or even correctness of vocabulary have to be considered. The list here is incomplete and the relations are rather simplified, evaluation can not be a focus in this paper. For more detailed evaluation strategies, please refer to Cole[19].

DISCUSSION

Continuous speech understanding systems are still years away before reasonably systems will be available commercially for clinical domains such as internal medicine. These speech understanding systems will enable smart user interfaces (UI). These UIs will not only know who is speaking to them, but also know the specific clinical domain of the speaker. In addition, the UI may learn and adapt not only from the speaker, but also from the patient data that is stored in the CPR. The user interface may be able to anticipate the health care provider's information needs and actions. Since such a system will know the meaning of clinical concepts and clinical processes, the system can become a true professional consultant or partner for the health care provider.

Clearly, the market is there, the computers are becoming faster and affordable to support natural speech recognition (most new computers will have built-in sound-board-capability), and the research is far enough along to be transformed into commercial products. In the long term, beyond three years, a big pay-off may be realized when intelligent computer-based patient records and voice understanding technologies are integrated.

In the health care market, we are at the beginning of a revolution. In the new health care environment, the financial risks of patient care are shifted from the insurance companies to the health care providers. When provider groups negotiate contracts with health insurance companies, they must know how much it costs them to cover episodes of care (e.g., a hip replacement, complicated pregnancy, HIV treatment). Thus the clinical data in the computer-based patient record is essential for their survival -- it must be collected. Simultaneously, the pressure to reduce cost for patient care services forces health care institutions to be more efficient in how they work. This includes more efficient data collection, more intelligent use of the collected information to avoid duplicate clinical tests and procedures. Voice understanding technology has the potential for increased efficiency because it brings the health care providers closer to the database, allowing them to manage the computer-based patient record directly and make better use of it. One beneficial side effect is better quality of patient care, which, in general, is associated with lower cost and certainly better patient satisfaction. The outlook for a voice understanding user interface in the health care industry has never been better.

References

1. Bergeron BP. Voice recognition: an enabling technology for modern health care? In: Cimino JJ, ed. AMIA Annu Fall Symp: Hanley & Belfus, Inc, 1996:802-806.
2. Isaacs E, Wulfman CE, Rohn JA, Lane CD, Fagan LM. Graphical access to medical expert systems: IV. Experiments to determine the role of spoken input. Methods Inf Med 1993;32(1):18-32.
3. Wingert F. Medical Linguistics: Automated Indexing into SNOMED. Critical Reviews in Medical Informatics 1 1988:333-403.
4. Haug PJ, Ranum PL, Frederick PR. Computerized extraction of coded findings from free text radiology reports. Radiology 1990(174):543-49.
5. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994;1(2):161-74.
6. Cimino JJ. Data storage and knowledge representation for clinical workstations. Int J Biomed Comput 1994(34):185-94.
7. Baud RH, Rassinoux A-M, Scherrer J-R. Natural Language Processing and Semantical Representation of Medical Texts. Meth. Inform. Med 1992;31:117-25.
8. van Ginneken AM. Structured Data Entry in ORCA: the Strength of two Models Combined. In: Cimino JJ, ed. Proc AMIA Annu Fall Symp. Washington, DC: Hanley&Belfus Philadelphia, 1996:797-701.
9. van Ginneken AM. The Structure of Data in Medical Records. In: van Bemmel JH, McCray AT, eds. Yearbook of Medical Informatics 1995, Stuttgart: Schattauer, 1995:61-70.
10. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A Model for Structured Data Entry Based on Explicit Descriptional Knowledge. Meth. Inform. Med. 1994;33:454-63.
11. Rector AL. Coordinating Taxonomics: Key To Re-usable Concept Representations. In: Barahona P, Stefanelli M, Wyatt J, eds. 5th Conf. on Artificial Intelligence in Medicine Europe (AIME '95). Pavia, Italy: Springer, Berlin, 1995:17-28.
12. Wagner JC, Solomon WD, Michel PA, et al. Multilingual natural language generation as part of a medical terminology server. In: Greenes RA, Peterson HE, Protti DJ, eds. Medinfo. Vancouver, Canada: North-Holland, 1995:100-104.
13. Birkmann C, Diedrich T, Ingenerf J, Rogers J, Moser W, Engelbrecht R. A Formal Model of Diabetological Terminology and Its Application for Data Entry. In: Pappas C, Maglaveras N, Scherrer J-R, eds. Proc MIE '97. Porto Carras, Greece: IOS Press, 1997:426-430.
14. Wormek A, Ingenerf J. Unterstützung der Befunderhebung in der Pathologie durch ein Spracherkennungssystem. In: Baur MP, Frimmers R, Blettner M, eds. Proc GMDS '96. Bonn: MMV Medizin, 1997:94 - 98.
15. Meijer GA, Baak JP, van Diest PJ, van Hattum AH, van der Linden HC, Koevoets JJ. Text processing by digital voice recognition. Anal Quant Cytol Histol 1996;18(4):261-266.
16. Teplitz C, Cipriani M, Dicostanzo DS, J. Automated Speech-Recognition Anatomic Pathology (ASAP) Reporting. Seminars in Diagnostic Pathology 1994;11(4):245-252.
17. Wulfman CE, Rua M, Lane CD, Shortliffe EH, Fagan LM. Graphical access to medical expert systems: V. Integration with continuous-speech recognition. Methods Inf Med 1993;32(1):33-46.
18. Gamma E, Helm R, Johnson R, Vlissides J. Design Patterns: Elements of Reusable Object-Oriented Software. Reading, Massachusetts: Addison-Wesley, 1995.
19. Cole RA, Mariani J, Uszkoreit H, Zaenen A, Zue V. Evaluation. Survey of the State of the Art in Human Language Technology. 1995:475 ff. http://www.cse.ogi.edu/CSLU